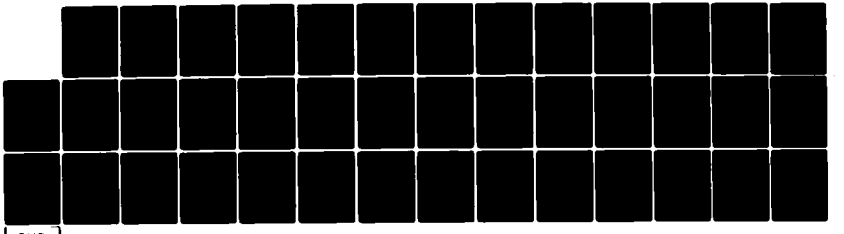
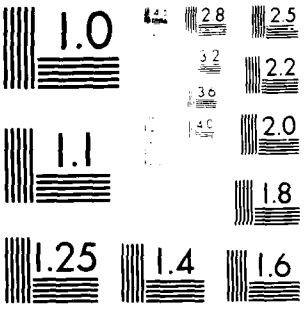
AD-A148 338 THE COMBINATORICS OF LOCAL CONSTRAINTS IN MODEL-BASED 1/1
RECOGNITION AND LOC..(U) MASSACHUSETTS INST OF TECH
CAMBRIDGE ARTIFICIAL INTELLIGENCE L.. W E GRIMSON
UNCLASSIFIED APR 84 AI-M-763 N00014-80-C-0505 F/G 9/4 NL



END
DATE FILMED
1-85
DTIC



MICROCOPY RESOLUTION TEST CHART

(2)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AIM 763	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Combinatorics of Local Constraints in Model-Based Recognition and Localization		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) W. Eric L. Grimson		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0505 N00014-82-K-0334
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE April 1984 13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Information Systems Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES None		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Object Recognition Model-Based Recognition Constraint Propagation Constrained Relaxation Combinatorial Analysis		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of recognizing what objects are where in the workspace of a robot can be cast as one of searching for a consistent matching between sensory data elements and equivalent model elements. In principle, this search space is enormous and to contain the potential combinatorial explosion, constraints between the data and model elements are needed. We derive a set of constraints for sparse sensory data that are applicable to a wide variety of sensors and examine their completeness and exhaustiveness. We then derive

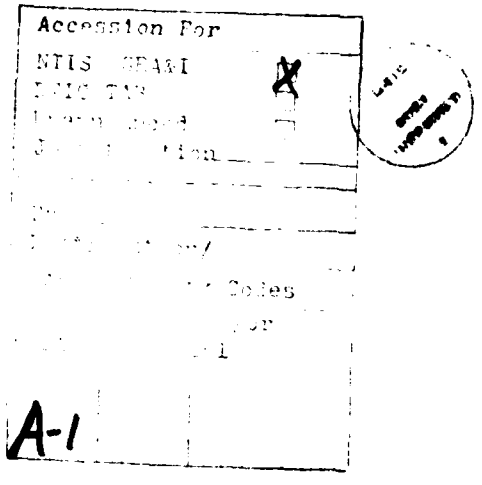
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 cont.

general theoretical bounds on the number of interpretations expected to be consistent with the data under the effects of local constraints. These bounds are applicable to many types of local constraints, other than the specific examples used here. For the case of sparse, noisy three-dimensional sensory data, explicit values for the bounds are computed and are shown to be consistent with empirical results obtained earlier in (Grimson and Lozano-Perez 1984). The results are used to demonstrate the graceful degradation of the recognition technique with the presence of noise in the data, and to predict the number of data points needed in general to uniquely determine the object being sensed.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A. I. Memo 763 April, 1984

The Combinatorics of Local Constraints in Model-Based Recognition and Localization From Sparse Data

W. Eric L. Grimson

Abstract. The problem of recognizing what objects are where in the workspace of a robot can be cast as one of searching for a consistent matching between sensory data elements and equivalent model elements. In principle, this search space is enormous and to contain the potential combinatorial explosion, constraints between the data and model elements are needed. We derive a set of constraints for sparse sensory data that are applicable to a wide variety of sensors and examine their completeness and exhaustiveness. We then derive general theoretical bounds on the number of interpretations expected to be consistent with the data under the effects of local constraints. These bounds are applicable to many types of local constraints, other than the specific examples used here. For the case of sparse, noisy three-dimensional sensory data, explicit values for the bounds are computed and are shown to be consistent with empirical results obtained earlier in [Grimson and Lozano-Pèrez 1984]. The results are used to demonstrate the graceful degradation of the recognition technique with the presence of noise in the data, and to predict the number of data points needed in general to uniquely determine the object being sensed.

Acknowledgments. This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's Artificial Intelligence research is provided in part by a grant from the System Development Foundation, and in part by the Advanced Research Projects Agency under Office of Naval Research contracts N00014-80-C 0505 and N00014 82 K-0334.

1. The Recognition and Localization Problem

A central characteristic of advanced applications in robotics is the presence of significant uncertainty about the identities and positions of objects in the workspace of the robot. In simplest terms, if a robot is to interact intelligently with its environment, it must know *what* objects are *where*. This normally necessitates sensing of the external environment as a means of obtaining the information needed to solve the recognition and localization problem. The process of sensing can be loosely divided into two stages: first, the measurements of properties of the objects in the environment, and second, the interpretation of those measurements. Since the sensory information could come from a variety of very different sources, for example, tactile, ranging, sonar or vision, both binary and grey-level, it is important to derive recognition and localization techniques that solve the interpretation stage of the sensing process with very few assumptions on the sensory measurements themselves. In this article, we assume only that the sensory data is characterized as sparse, noisy measurements of the local geometry of a small patch of the object's surface, for example, the position and orientation of a small planar patch of the surface in some coordinate frame defined relative to the sensor.

Given these simple data elements derived from the sensory data, the problem of model-based recognition and localization essentially can be considered as one of searching a space for a consistent matching between these data elements on the one hand, and model elements representing known objects on the other hand. Since the data elements are assumed to approximate the local geometry of a small planar patch of the surface, initially we assume that the objects can be modeled as polyhedra having up to six degrees of freedom relative to the sensor. The goal is then to define a matching process whereby the space of possible interpretations of the sensory data can be searched for a consistent matching of sensory data to model elements.

Since, in general, the search space is far too large to explicitly explore, the key to the problem is to derive constraints on the data that will efficiently restrict the portions of the search space that must be explored. In this paper, we present a set of general criteria on these constraints and derive a specific set of such constraints for the case of geometric sensor measurements. Using this particular example, we show that these constraints are complete both for the simpler case of three degrees of freedom (isolated objects in stable positions) and for the general case of six degrees of freedom. We also show that the constraints are exhaustive for the three degree of freedom case, but not for six degrees of freedom. By exhaustive, we mean that modulo errors in the measurement process, the measured constraints are sufficient to completely determine the relative configuration of the sensory data.

The main result is establishing theoretical bounds on the effectiveness of local constraints in controlling the combinatorics of the search process. We stress that the results obtained hold in large part independent of the specific constraints used. For the particular constraints derived here, we also compute explicit values for the predicted bounds and compare them with empirical evidence derived earlier

in [Grimson and Lozano-Pèrez 1984]. Finally, several predictions of the theory are discussed, including the degradation of the technique with increased error, and the number of sensory points generally needed to guarantee a unique interpretation of the data.

1.1. The Basic Problem

The recognition and localization of objects from sensory data is a central problem of most advanced robotics situations. It is usually convenient to pose the problem as one of search, that is, given a set of known models, we identify and locate the particular object that we are sensing by searching a large space of possible solutions until we find one (or all solutions) that matches the information available to us from the sensors. One of the main difficulties with the problem, true of most search problems, is that the space of possible solutions is usually extremely large, and one seeks methods that will effectively reduce the portions of the search space that must be explicitly explored. The problem is further compounded in the case considered here by the fact that the sensory data against which a match is sought are typically inaccurate, so that the matching process must be tolerant to errors in the data.

The critical issue in searching for a match between sensory data and object models is controlling the potential combinatorial explosion of the search. There have been a wide variety of techniques applied to the recognition problem, all attempting in some manner to control this explosion. We can distinguish three general classes of approaches, although these distinctions are not hard and fast, of course. The three general classes are (i) matching complete descriptions of the object obtained from the sensory data to complete model descriptions, (ii) matching partial descriptions of the data to partial descriptions of the model, and (iii) matching partial descriptions of the data to complete descriptions of the model.

The basic idea behind matching complete descriptions to complete descriptions is to reduce the combinatorics by computing compact representations of a dense set of sensory data and comparing this representation to a similar object model. In this manner, the matching process is constrained to a small number of components and the combinatorics is significantly reduced. Examples include generalized cylinder representations (see for example, Nevatia 1974; Nevatia and Binford 1977; Marr and Nishihara 1978; Brooks 1981), and extended Gaussian images (see for example, Brou 1983; Horn 1983; Horn and Ikeuchi 1983; Ikeuchi 1983). Besides the combinatorial advantage of comparing them, complete descriptions have several other potential advantages. First, by building the representation from a large, dense set of sensory data, it is likely to be insensitive to errors in the individual data elements. Second, even if only parts of the object are available to the sensor, the use of complete representations makes it likely that recognition can still proceed on the partial data. Of course, to some extent, the reduction in computational cost achieved by matching compact descriptions is offset by the additional cost of processing the raw sensory data to obtain those representations. We also note that not all global representations share the advantages cited above. For example, one of the earliest

techniques for recognition uses global properties of binary images, such as area, perimeter, elongation and Euler number, an approach common to many commercial systems (see for example, Bausch and Lomb 1976; Gleason and Agin 1979; Machine Intelligence Corporation 1980; Reinhold and Vanderbrug 1980). These particular parameters do not extend well to overlapping or occluded parts, and thus such techniques do not demonstrate the same versatility as the previous ones.

A second approach to the problem is to use input tokens (frequently called features) to the matching process that are very distinctive. Typically, if one can obtain distinctive features, then the search reduces to a straightforward depth first exploration of an interpretation tree, with very little backtracking involved. For example, if I wanted to recognize a soft drink can from visual data, I could process the image to obtain the UPC bar code, which would uniquely identify the type of can. Moreover, knowing the position of the UPC code on the can and in the image would allow me to determine the position and attitude of the can in the scene. Of course, not all features will be as distinctive as a UPC code. Simpler examples might include corners, holes, notches and other local features. The idea, however, is that very few such distinctive features should be needed to identify the object, and the search space can be effectively collapsed. Examples of techniques in this vein include the use of a few extended features [Perkins 1978; Ballard 1981], or the use of one feature as a focus, with the search restricted to a few nearby features [Tsuji and Nakamura 1975; Holland 1976; Sugihara 1979; Bolles and Cain 1982; Bolles, Horaud and Hannah 1983]

There are several drawbacks to such distinctive feature approaches, although they also have many strong points. First, while the cost of the search process has been greatly reduced in this case, it is usually at the expense of the processing of the sensory data. In other words, the unique features required to straightforwardly identify the object are usually not directly provided by the sensors. Rather, the data provided by the sensors is typically point data, for example, local measurements of the position and/or orientation of points on a surface. Each such local point measurement is clearly not very distinctive, and to create distinct collections or features requires additional preprocessing of the data. In the example of the UPC code, the visual input must be processed to extract the code from the rest of the data. Thus, feature driven recognition techniques usually reduce the computational expense of searching a space of solutions at the increased cost of computing the tokens to be matched between sensory data and model elements.

Not only is there additional computational cost for computing appropriate features, but the computation may also be sensitive to sensor errors. In the example of the UPC code, if the imaging device is out of focus, causing the image of the bar code to blur significantly, the recognition process may no longer succeed. Our preference is for recognition techniques that degrade gracefully with noise, rather than suddenly collapsing under the influence of sensor error.

Clearly these problems do not rule out feature driven recognition schemes, as is evidenced in our example of the soft drink can, by the proliferation of automatic check-out counters in supermarkets. A more critical drawback concerns the density,

or rather, sparsity of such features in the sensory data. By definition, features well suited for recognition and localization must be sparse on the object, since otherwise there would be multiple interpretations of the attitude of the object relative to the sensor. While this eases the recognition task, it also reduces the situations in which the process can be applied. In particular, it requires that the initial sensory data be dense, in order to have a reasonable expectation that the feature extraction stage will actually find a set of useful features from the data. This requirement on dense sensing modalities rules out some types of sensors, such as tactile, that are inherently sparse in nature. If possible, we would prefer a recognition technique that is sensor-independent.

The sparsity of features is also a problem when dealing with occlusion. If an object is presented to the sensor in isolation, then large portions of its surface will be visible and the probability of detecting appropriate features is high. If the object is partially occluded by other objects, however, this may not be true. In our example of a soft drink can, if some other object occludes the UPC bar code from the sensor, we will not be able to recognize the can. Note that this may occur even though virtually all the rest of the can is visible to the sensor. This follows in general from the sparsity of useful features. If possible, we would like to use recognition techniques that are not dependent on particular localized features of the objects, and are thus, still capable to performing recognition and localization on partially occluded objects.

In summary, our main concern with distinctive feature techniques is that because they are matching partial descriptions to partial descriptions, it may occasionally be difficult to guarantee the computability of such descriptions from the raw data. This implies that the inherent sparseness of such features on an object may cause problems for sparse sensors, such as tactile sensors, and may cause problems in situations involving occlusion.

A third approach to the recognition and localization problem is to use the local point measurements available from the sensors as the basic matching tokens. Of course, in some sense these are also features, but they do not suffer the same problems due to sparseness, since they are dense on the object. Since these data elements are very simple, taken individually they are not likely to uniquely identify the object being sensed. Thus the search part of the process becomes much more critical, and we will need to use local constraints between such point measurements to restrict the search space. While the size of the search space explored in this case will be larger than in the feature based methods, the expectation is that the unit cost of searching the space can be reduced significantly enough to overcome the number of additional elements tested in the search. Representative examples of such schemes include Faugeras and Hebert [1983], Gaston and Lozano-Pérez [1984], Grimson and Lozano-Pérez [1984] and Stockman and Esteva [1984].

The key distinction between schemes that match partial descriptions using low level sensory measurements and schemes that match partial descriptions based on distinctive features lies in the computability of the sensory data descriptions. Simple, low-level sensor measurements are likely to be dense over the object. As

a consequence, recognition schemes based on such simple measurements should be applicable to sparse sensors, and should be insensitive to problems of occlusion and sensor error, since an input description can always be obtained and matched to the model. As the sensor features to be matched become more distinctive, and hence sparser on the object, the probability that they may not be detected by the sensor, either due to the characteristics of the sensor or to problems of occlusion rises. In this paper, we will explore the use of recognition schemes that use low level sensor primitives that can be computed over the entire object. We will, however, consider using only a sparse set of such measurements, in order to keep the combinatorics of the search process reasonably controlled.

1.2. Assumptions and Approach

As a consequence of this discussion, we will assume that the basic sensory data available consists of local estimates of three-dimensional positions and orientations of small patches of the object surface. In this case, we can make very simple assumptions about the elements of the object models needed for matching. In particular, since the data elements are measuring the local geometry of small patches of the object surface, we assume that the object models are also constructed of small local patches. Thus, our two assumptions about the elements to be matched between sensory input and object models are:

1. The objects are all modeled as polyhedra having six degrees of freedom.
2. The sensory data available to the process include positions of points on the object, to within some volume of error, and surface orientations at those points, to within some cone of error.

The basic approach to the problem is to determine the set of positions and orientations of an object that are consistent with this sensed data. If there are no consistent positions and orientations, then the object can be excluded from the set of possible objects. The elements to be matched are thus simple local patches of a surface.

The technique proceeds in two steps:

1. *Generate Feasible Interpretations:* A set of feasible interpretations of the sense data is constructed. Interpretations consist of pairings of each sensed point with some object surface on one of the known objects. Interpretations inconsistent with local constraints, derived from the model, on the sense data are discarded.
2. *Model Test:* The feasible interpretations are tested for consistency with surface equations obtained from the object models. An interpretation is legal if it is possible to solve for a rotation and translation that would place each sense point on an object surface. The sensed point must lie *inside* the object face, not just on the surface defined by the equation.

There are several possible methods of actually searching for consistent matches. For example, Grimson and Lozano-Pérez [1984] chose to structure the search as the generation and exploration of an interpretation tree. That is, starting at a root

node, we construct a tree in a depth first fashion, assigning data points to model faces. At the first level of the tree, we consider assigning the first data point to all possible faces, at the next level, we assign the second data point to all possible faces, and so on.

Clearly, the first step is the key to this process. The number of possible interpretations given s sensed points and n surfaces is n^s. Therefore, it is not feasible to explore the entire search space in order to apply a model test to all possible interpretations. Moreover, since the computation of coordinate frame transformations tends to be expensive, we want to apply this part of the technique only as needed. The goal of the recognition algorithm is thus to exploit local constraints on the sensed data so as to minimize the number of interpretations that need testing, while keeping the computational cost of each constraint small. In the case of the interpretation tree, we need constraints between the data elements and the model elements that will allow us to remove entire subtrees from consideration without explicitly having to search those subtrees.

In searching for appropriate constraints to apply to the generation stage, several criteria are appropriate.

1. The constraints should be coordinate frame independent. That is, we would like to derive constraints that remove large portions of the search space, independent of the particular orientation of the object. This suggests that the constraints should embody restrictions due to the shape of the object, and not due to the specifics of the sensing geometry.
2. The constraints should be simple and have low computational cost.
3. The constraints should, at the same time, be as powerful as possible in the sense of removing large portions of the overall search space.
4. The constraints should degrade gracefully in the presence of error in the sensory measurements.
5. The constraints should be independent of the specifics of the sensor from which the data came, so that they will apply equally to different sensing modalities.

These constraints are very similar to those suggested by Marr and Nishihara [1978] (see also Marr [1982]).

2. A Specific Set of Local Constraints

We begin our analysis by first deriving a set of coordinate frame independent constraints, that were first presented in [Grimson and Lozano-Pérez 1984]. The first question posed is what types of coordinate-frame-independent constraints are possible, given that the sensory data is characterized as sparse points, each consisting of a position measurement and a unit surface normal (see Figure 1). Clearly a single data point provides no constraint on the possible faces from the model that could consistently be assigned to it. Thus, we look at pairs of sensory

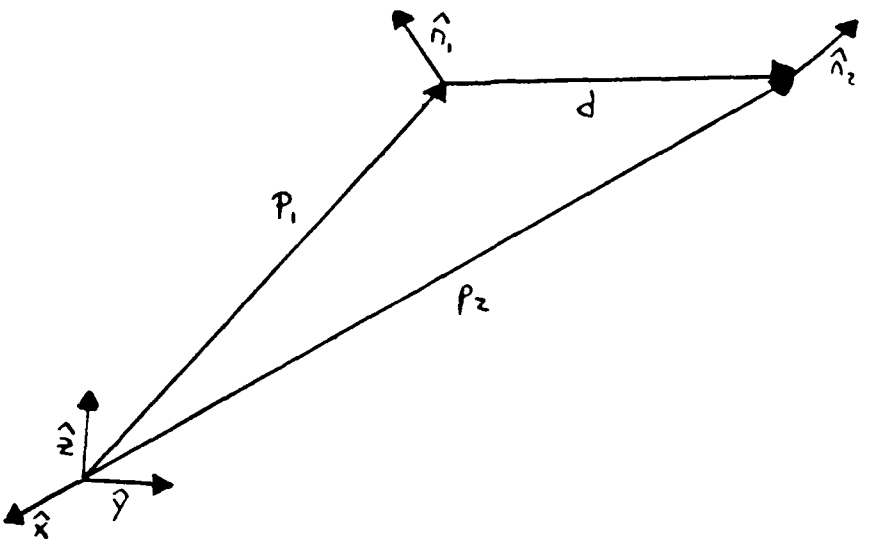


Figure 1. The constraints between pairs of sensory points.

points, and the basic information available from these points consists of a pair of unit normals, as well as the vector separating their bases, as shown in Figure 1.

One way to get coordinate-frame-independent constraints is to construct a local coordinate frame relative to the configuration itself. Thus, we can use each of the two unit normals as axes of the local coordinate frame. In two dimensions, these define a local system, except in the degenerate case of the unit normals being (anti-)parallel. In three dimensions, the third component of the local coordinate frame can be taken as the unit vector in the direction of the cross product of the first two basis vectors. Given such a local basis, clearly one set of coordinate-frame-independent measurements provided by the configuration of Figure 1 is the components of the separation vector along each of the basis directions. (Note that the use of the distance and two of the components is equivalent, up to a possible sign ambiguity, to using the three components of the vector.) Additionally, the angle between the two basis vectors is also specific only to the local coordinate frame. More formally, if the unit vectors are denoted by $\mathbf{n}_1, \mathbf{n}_2$ and the vector separating the two points is $\mathbf{d}$, then one set of coordinate-frame-independent measurements of this configuration is

$$\mathbf{n}_1 \cdot \mathbf{n}_2$$
$$\mathbf{d} \cdot \mathbf{n}_1$$
$$\mathbf{d} \cdot \mathbf{n}_2$$
$$\mathbf{d} \cdot \mathbf{u}$$

where $\mathbf{u}$ is a unit vector in the direction of $\mathbf{n}_1 \times \mathbf{n}_2$. (Note that these measurements are isomorphic to the set used in [Grimson and Lozano-Pérez 1984].)

These are coordinate-frame-independent measurements on the configuration defined by a pair of sensory points. To turn them into constraints on the search process, we must map them into equivalent measurements on the model elements. Since each object is modeled as a complex polyhedra, the mapping is fairly straightforward. For example, consider the first measurement, $\mathbf{n}_1 \cdot \mathbf{n}_2$. In order for these two sensory points to be consistent with a pair of faces on an object, the dot product of the corresponding face normals must agree with this measurement. That is, by precomputing the angles between all pairs of faces on an object, this sensory measurement can be used to constrain the search for a consistent interpretation. In particular, if a pair of points is inconsistent with a particular assignment of faces to those points, the entire subtree of the interpretation tree lying below the node corresponding to that assignment can be ignored, thereby reducing the amount of searching required. Similar constraints can be derived for the components of the separation vector in the directions of the unit normals. That is, for each pair of faces in the model, one can precompute the range of values of the component of a vector in the direction of one of the face normals as that vector assumes all possible positions having one endpoint on the first face and the other endpoint on the second. Again, for an assignment of sensory points to faces to be consistent, it must be the case that the coordinate-frame-independent sensory measurement must agree with the precomputed model values. Thus, we have defined one possible matching process between sensory measurements and models, similar to that presented in [Grimson and Lozano-Pérez 84].

3. The Constraints Satisfy Our Criteria

We began our theoretical investigation by establishing that the constraints derived above meet the criteria of simplicity, coordinate-frame-independence, completeness and exhaustiveness. We will then consider their combinatorial power and their degradation with error.

3.1. Coordinate-Frame-Independence

Clearly by the specifics of the derivation, the constraints are coordinate-frame independent. Moreover, they also satisfy our requirement of simplicity. Obtaining the sensory half of the constraint is straightforward, and the model half of the constraints can be precomputed directly from the model. The matching process itself then becomes a simple table-lookup process, which also satisfies the notion of simplicity and computational ease.

3.2. Completeness of the Constraints

While the constraints derived above meet our basic criterion, it is important to also demonstrate that they form a complete set, that is, there are no other

independent coordinate frame free constraints between pairs of sensory points that are not already incorporated into these particular constraints. This can be easily established by the following argument. Consider the configuration illustrated in Figure 1. Suppose we construct some arbitrary global coordinate frame, having three rotational degrees of freedom ϕ, θ, ψ and three translational degrees of freedom, given by the components of the vector $\mathbf{p}_1$, relative to the described configuration. To completely describe the configuration shown will require a total of 10 equations, three to specify the base position $\mathbf{p}_1$, three more to specify the separation vector $\mathbf{d}$, and two each to specify the relative attitudes of $\mathbf{n}_1, \mathbf{n}_2$. These ten equations are explicit functions of several parameters, including the three angular and three translational degrees of freedom described above. Thus, to reduce this set of 10 constraints to a set of coordinate-frame-independent constraints, we must resolve the set of equations to remove the explicit dependence on the parameters related to the specific choice of global coordinate frame. Clearly this will require at least six equations, and thus there are at most four coordinate-frame-independent constraints given by this particular pairwise configuration of sensory points.

In the simpler case of two dimensions, we find that there are three coordinate-frame-independent constraints. Thus, we see that in both cases, the constraints outlined above are complete.

3.3. Exhaustiveness

While the above analysis indicates that the set of constraints is complete, in that there are no additional independent constraints possible, we can also examine the exhaustiveness of these constraints. In particular, we can ask the following question. Given the configuration of Figure 1, suppose we know the position and orientation of one of the endpoints and its orientation vector in some coordinate frame. Does the information provided by the local constraints of the above section uniquely determine the position and orientation of the second point uniquely? In other words, ignoring issues of error, how well do the local measurements restrict the possible interpretations of the data? By counting degrees of freedom, we would expect the second point to be uniquely determined for the two dimensional case, but not for the three dimensional one. We now proceed to establish this claim.

3.3.1. Two Dimensions

We first consider the case of two dimensions. Given one sensory point, consisting of a position vector and an associated unit normal vector, we construct a relative coordinate system with origin at the end of the position vector and with $-y$ axis oriented along the unit normal. Thus, a second sensory point can be characterized in this space by a unit vector $\mathbf{n}$ offset by some other vector $\mathbf{p}$, which we represent by the ordered pair

$$(\mathbf{p}, \mathbf{n}).$$

To determine the exhaustiveness of the constraints, we need to show that the pair $(\mathbf{p}, \mathbf{n})$ is uniquely determined by the following measurable parameters:

1. The angle θ between the unit normals.

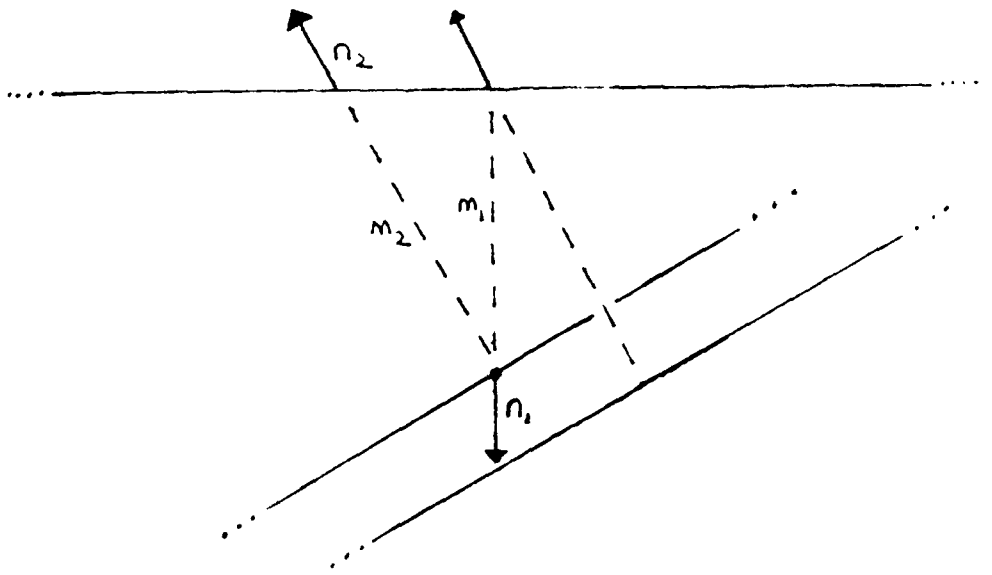


Figure 2. The completeness of the constraints in two dimensions.

2. The components m_1 and m_2 of the vector p relative to the two unit normals respectively.

By definition of the local coordinate system, $\mathbf{n}_1 = (0, -1)$, and since the angle θ can be determined from the sensory measurements, the second unit normal is also completely determined, $n_2 = (\sin\theta, -\cos\theta)$.

To begin, we ignore the special degenerate case of

$$(\mathbf{n}_1 \cdot \mathbf{n}_2)^2 = 1.$$

In this case the vector p can be represented by

$$\mathbf{p} = \alpha\mathbf{n}_1 + \beta\mathbf{n}_2$$

where α and β are parameters to be determined. By taking the dot product of p with each of the unit normal vectors, we obtain the following system of equations:

$$\alpha + \beta\cos\theta = m_1$$
$$\alpha\cos\theta + \beta = m_2.$$

Since $\cos\theta \neq \pm 1$, we can solve this system for α and β, yielding

$$\mathbf{p} = \frac{m_1 - \cos\theta m_2}{\sin^2\theta}\mathbf{n}_1 + \frac{m_2 - \cos\theta m_1}{\sin^2\theta}\mathbf{n}_2.$$

Thus, the ordered pair $(\mathbf{p}, n)$ is completely determined by the measured values.

Geometrically, we note that the component m_1 constrains p to lie on the line $y = -m_1$ as shown in Figure 2. Since the orientation of the second unit normal is known, the final constraint comes from the component of the vector between the two sensed points in the direction of this normal, in particular, that $-\mathbf{p} \cdot \mathbf{n} = m_2$.

In Figure 2, this corresponds to finding the ray perpendicular to $\mathbf{n}$ at a specified distance that contains the origin. This ray completely determines the geometry.

In the degenerate case of the two unit normals being (anti-)parallel, we can use the magnitude of the distance vector $d = \sqrt{\mathbf{p} \cdot \mathbf{p}}$ as an additional constraint in place of the now redundant constraint $\mathbf{p} \cdot \mathbf{n}_2 = m_2$. In this case, the distance constrains $\mathbf{p}$ to lie on a circle of radius d. Combining this with the constraint given by the component m_1 restricts $\mathbf{p}$ to one of two positions, and there are two solutions

$$(\mathbf{p}, \mathbf{n}) = \left(\left[\pm\sqrt{d^2 - m_1^2}, -m_1 \right], \left[0, -\frac{m_2}{m_1} \right] \right)$$

where $|m_1| = |m_2|$.

Thus, over the set of all possible relative orientations of faces, the constraints for the two dimensional case are almost always complete. This is illustrated in Figure 2.

3.3.2. Three Dimensions

We can consider a similar analysis in the three dimensional case. As before, we construct a relative coordinate system with origin at the first position vector and with the $-z$ axis point along the surface normal. The x and y axes can be oriented arbitrarily. Here, we have a five degree of freedom problem, since we must determine the ordered pair

$$(\mathbf{p}, \mathbf{n})$$

where $\mathbf{p}$ has three degrees of freedom

$$\mathbf{p} = (x, y, z)$$

and where $\mathbf{n}$ has two degrees of freedom

$$\mathbf{n} = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi) \qquad -\pi < \theta \leq \pi, 0 \leq \phi < \pi.$$

The constraints in this case are:

1. The dot product between the two surface normals, ϵ.
2. The components m_1 and m_2 of the vector $\mathbf{p}$ relative to the two unit normals, $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively.
3. The component m_3 of the vector $\mathbf{p}$ relative to the cross product of the two unit normals, $\mathbf{n}_1 \times \mathbf{n}_2$.

As in the two dimensional case, we exclude the degenerate case of

$$(\mathbf{n}_1 \cdot \mathbf{n}_2)^2 = 1.$$

A similar analysis holds in that case, using the distance constraint $d = \sqrt{\mathbf{p} \cdot \mathbf{p}}$ instead.

The constraints essentially supply four equations in five unknowns, x, y, z, θ, ϕ that we wish to resolve. The equations are given by

$$\begin{aligned} -z &= m_1 \\ \sin\phi(x\cos\theta + y\sin\theta) + z\cos\phi &= m_2 \\ \sin\phi(x\sin\theta - y\cos\theta) &= m_3 \\ -\cos\phi &= \epsilon. \end{aligned}$$

Straightforward algebraic manipulation yields the following parameterized solution:

$$\begin{aligned} x(\theta) &= \frac{1}{\sin\phi}(m_3\sin\theta + [m_2 - \epsilon m_1]\cos\theta) \\ y(\theta) &= \frac{1}{\sin\phi}(-m_3\cos\theta + [m_2 - \epsilon m_1]\sin\theta) \\ z &= -m_1 \\ \mathrm{n}(\theta) &= (\sin\phi\cos\theta, \sin\phi\sin\theta, -\epsilon) \end{aligned}$$

where $\sin\phi$ can take on one of two values,

$$\sin\phi = \pm\sqrt{1-\epsilon^2}.$$

In more geometric terms, this particular resolution of the equations restricts p to lie on a circle, and for each point on that circle, there are two possible values for the orientation of the surface normal n. Thus, we see that these pairwise constraints alone do not completely determine the configuration of a pair of sensory points, even in the general case.

4. A General Theoretical Basis

The key theoretical issue still to be settled is the combinatorial power of the local constraints in reducing the number of consistent hypotheses. In this section, we develop a theoretical basis for analyzing the combinatorics of the recognition process. We will see that while worst case bounds on the number of interpretations consistent with sparse data tend to be very weak, expected case bounds turn out to be very strong, and are in fact supported by empirical evidence. We will begin our study by first considering the three degree of freedom case. We will then extend the analysis is a straightforward manner to the full six degree of freedom case. The first question we consider is that of deriving theoretical bounds on the efficiency of the constraints in restricting the search space. We will derive bounds that are independent of the specific constraints described above, and then show how well these particular constraints perform in restricting the search space.

4.1. Relative Configuration Space (RC-space)

Our general method of analysis can be summarized in the following manner. Since all of the constraints available to the recognition process are pairwise constraints, we have no information to apply to the assignment of a face to the first sense point, P_1. Hence, we will arbitrarily assign a face f_i to this point, compute a bound on the number of interpretations consistent with that choice, and then sum over all such assignments of f_i. To obtain a bound on the number of interpretations,

we will use the notion of a relative configuration space (RC-space). Throughout the analysis we will consider two separate cases. The first is one in which the objects are non-overlapping and lie in stable positions. This case is essentially a two dimensional one, and the objects have three degrees of freedom relative to their models, two translational and one rotational. The second case is one in which the objects are arbitrarily oriented in space. This is a three dimensional problem and the objects here have six degrees of freedom relative to their models, three translational and three rotational.

In the two dimensional case, we define an RC-space in the following manner. Each face of an object is defined by a two-dimensional position (given by the position of the midpoint of the face) and an orientation (given by the orientation of the normal of the face). To ease the analysis, we will assume that all of the edges of the object (or polygon) have equal length ℓ. For a three dimensional object, we assume that all of the faces are squares of side ℓ.

Given a face f_i assigned to the first point P_1, we can define a three dimensional coordinate system relative to this face, consisting of two coordinates of spatial extent and one of angular extent. The origin of the two spatial dimensions is set at the midpoint of the base face (f_i) (with the face extending along the x axis and with the normal of the face pointing in the $-y$ direction) and the origin of the angular dimension is set to the orientation of the normal of the base face. Thus we have defined a configuration space relative to the orientation of a particular face. The position of another face is completely specified by the position and orientation of its midpoint in this RC-space.

In the three dimensional case, we construct a 5 dimensional relative configuration space, based on the first face f_i. This RC-space has three positional dimensions and two orientation dimensions (since only two angles are needed to specify the orientation of a unit vector). Here, the origin of the spatial components is defined to be the midpoint of the base face, with the surface normal pointing in the $-z$ direction. The edges of the face are aligned with the x and y axes. Thus, the position of any other face is given by the position of its midpoint in this RC-space. The rotational components of a face are defined by two different angles, $-\pi < \theta \leq \pi$ and $0 \leq \phi < \pi$, where ϕ describes the elevation of the unit normal relative to the $-z$ axis, and θ describes the orientation of that unit normal in the $x - y$ plane.

To obtain bounds on the effectiveness of local constraints in pruning the search space of feasible interpretations, we need to map both the models and the sensory information into this RC-space. By enumerating the intersection of these two mappings, we will be able to analyze the combinatorial efficiency of the constraints.

Clearly, given a base face for the RC-space, each additional face of the model is represented by the position of its midpoint in this space, and that is given by the spatial offset of its midpoint relative to the midpoint of the base face, and the angular variation of its normal, relative to the normal of the base face. Thus, each face projects to a point in this RC-space, and the entire object is given by a scattered collection of such points.

Consider now what the sensory constraints tell us about the positions of possible faces in this RC-space. With no sensory constraint, each face of the object could lie anywhere within the RC-space. The sensory constraints on distance and relative angle, however, will restrict the set of possible faces to lie within a reduced volume in RC-space. Thus, given the assignment of the first point to face f_i, the second face can be restricted to lie within some finite volume of RC-space. Clearly, this should generally place a restriction on the number of faces consistent with the sensory data, and our goal is to obtain bounds on this restriction, by considering the characteristics of this restricted volume in RC-space.

In more detail, we consider what the effects of the sensory constraints are in the three degree of freedom case. In particular, given a second sense point P_2, we know the following facts.

1. Distance. The measured distance between the two sense points clearly restricts the position components of the second face to lie within a restricted area. This follows from the observation that if the base of the vector connecting the two sense points must lie somewhere on the first face, and the length of the vector is given by a measured distance d, which is accurate to within a range ϵ, then the set of positions in which the midpoint of the second face can lie is restricted to lie within an area specified by d and ϵ. (We will see later one method for analytically specifying this area.)
2. Angle of the vector. The measured components of the vector between sensed points relative to the measured surface normals define the angle between those vectors θ. This angle is subject to a range of values, which in this case is a function of d, ϵ and γ, the bound on errors in measuring surface normals.
3. Angle of the normal. The angle between the surface normals at the first and second faces can be restricted to some measured value ψ plus an error range defined by γ.

As a consequence of these constraints, we see that if the first point lies somewhere on the base face f_i, then the second point must lie within some volume

$$\mathbf{V}(d, \epsilon, \ell, \theta, \psi, \gamma)$$

of RC-space. Note that three of the parameters defining this volume are measured values and three are global parameters set by the object and the error sensitivities of the measuring device. To reflect this, we rewrite the volume of RC-space as

$$\mathbf{V}_{\epsilon,\ell,\gamma}(d, \theta, \psi).$$

Note that this symbol represents a region in a three-dimensional RC-space. We will use

$$v_{\epsilon,\ell,\gamma}(d, \theta, \psi)$$

to denote the magnitude of this volume.

Given that face f_i is initially assigned to the first sense point, we let

$$\rho_i(x, y, \phi)$$

denote the distribution of faces relative to this assignment. Thus, ρ_i is a sum of n delta functions, each of which reflects the configuration of a face relative to f_i, and

$$\int_x \int_y \int_{\phi=0}^{2\pi} \rho_i(x, y, \phi)\, dx\, dy\, d\phi = n.$$

If we let $d_{ij}, \psi_{ij}, \theta_{ij}$ denote the sensed measurements between points P_i and P_j, then the number of interpretations consistent with $k+1$ points of sensory data, given face f_i assigned to P_1 is bounded by

$$\prod_{j=2}^{k+1} \int\int\int_{\mathbf{V}_{\epsilon,\epsilon,\gamma}(d_{1j},\theta_{1j},\psi_{1j})} \rho_i(x, y, \phi)\, dx\, dy\, d\phi$$

and the total number of interpretations, over all possible assignments of the first point is bounded by

$$\sum_{i=1}^{n} \prod_{j=2}^{k+1} \int\int\int_{\mathbf{V}_{\epsilon,\epsilon,\gamma}(d_{1j},\theta_{1j},\psi_{1j})} \rho_i(x, y, \phi)\, dx\, dy\, d\phi. \tag{1}$$

Clearly, in the worst case, this is bounded above by $n \cdot n^k = n^{k+1}$, as would be expected. We now consider means for obtaining better bounds.

Note that the expression above only considers the effects of the sensory constraints between the first and the j^{th} point, and ignores the effects of intermediate points in constraining the possible values for the j^{th} face. Clearly, a better bound would be given by

$$\sum_{i=1}^{n} \prod_{j=2}^{k+1} \int\int\int_{\bigcap_{m=1}^{j-1} \mathbf{V}_{\epsilon,\epsilon,\gamma}(d_{mj},\theta_{mj},\psi_{mj})} \rho_i(x, y, \phi)\, dx\, dy\, d\phi. \tag{2}$$

Even here, the worst case behavior is still n^{k+1}.

There are several ways we could obtain tighter bounds. One is to note that for most realistic objects, there is a minimum distance between different faces, and thus the density of faces in RC-space has an upper limit. This could be used to place bounds on the expressions derived above. The second method is to seek expected bounds, that is, bounds that will hold in general, so that the class of objects for which the bound is exceeded is small, and hopefully, has measure zero, yielding an almost everywhere bound. This relies in some sense on a type of limiting behavior assumption as follows. Suppose we consider k different polygons, each with n faces and diameter D. Then in the limit as k tends to ∞, the distribution of the nk faces tends towards a uniform distribution. If we then average over k, we can assert that the expected distribution of faces is also uniform, over an area defined by the diameter of the object, D.

Thus, we let v_T denote the magnitude of the total usable volume of RC-space, given by

$$v_T = \left[\pi D^2\right] 2\pi.$$

The key point of this assumption of uniform distribution is that the integral of the distribution function will depend only on the magnitude of the swept volume, and not on its specific position in RC-space,

$$\int\int\int_{\mathbf{V}_{\epsilon,\ell,\gamma}(d_{mj},\theta_{mj},\psi_{mj})} \rho_i(x,y,\phi)\,dx\,dy\,d\phi = n\frac{v_{\epsilon,\ell,\gamma}(d_{mj},\theta_{mj},\psi_{mj})}{v_T}$$

Thus, if we can bound the magnitude of the volume

$$v_{\epsilon,\ell,\gamma}(d_{mj},\theta_{mj},\psi_{mj})$$

independent of the specific sensory measurements, say,

$$v'_{\epsilon,\ell,\gamma} = \sup_{d,\theta,\psi}\left[v_{\epsilon,\ell,\gamma}(d,\theta,\psi)\right],$$

then equation (1) leads to the expected bound of

$$n\left[n\frac{v'_{\epsilon,\ell,\gamma}}{v_T}\right]^k. \qquad (3)$$

This is still a relatively weak expected bound, since it is basically of the form n^{k+1}. Unless $nc \leq 1$ this bound will increase with increasing k, even for fixed c. Thus, if the volume of RC-space swept out by a single set of pairwise constraints is less than $\frac{1}{n}$ times the total available volume, the expected number of interpretations will decrease with increasing sensory information.

We should be able to obtain a tighter bound by using equation (2). This equation reflects that fact that while the constraints between the first and third sensory points yield a volume $v_{\epsilon,\ell,\gamma}$ for possible choices for the third face of an interpretation, the constraints between the second and third sensory points will in general yield a second volume of possible choices for the third face. In general, the intersection of these two volumes will be a smaller volume of consistent choices for the third face. This observation, of course, extends for all further faces in an interpretation. In order to provide bounds on the magnitude of this intersected volume, independent of the position and orientation of the additional sensory points relative to the first one, we consider the following technique.

First, we decouple the spatial and angular components of the RC-volume,

$$\mathbf{V}_{\epsilon,\ell,\gamma} = \mathbf{V}^s_{\epsilon,\ell,\gamma} \bigotimes \mathbf{V}^a_{\epsilon,\ell,\gamma}.$$

We let m denote the maximum separation of points lying in the spatial portion of the volume, $\mathbf{V}^s$. Then we can define $\mathbf{C}_{\epsilon,\ell,\gamma}$ to be the cube (square in the case of two spatial dimensions) of side m centered about the center of mass of $\mathbf{V}^s_{\epsilon,\ell,\gamma}$. Finally, we let

$$\mathbf{V}^*_{\epsilon,\ell,\gamma} = \mathbf{C} \bigotimes \mathbf{V}^a_{\epsilon,\ell,\gamma}.$$

In intuitive terms, $\mathbf{V}^*$ represents a larger volume of RC-space that contains the volume of possible configurations for a sensory point, independent of the specific orientation of previous sensory points. As a consequence,

$$\bigcap_{m=1}^{j-1} \mathbf{V}_{\epsilon,\ell,\gamma}(d_{mj}, \theta_{mj}, \psi_{mj}) \subseteq \bigcap_{m=1}^{j-1} \mathbf{V}^{*}_{\epsilon,\ell,\gamma}(d_{mj}, \theta_{mj}, \psi_{mj}).$$

In order to bound the magnitude of this new volume, we rely on the following two lemmas.

Lemma 1. Consider a parameter range of extent a and a second parameter range of extent b. If the two ranges must overlap and the ranges are otherwise independent, the expected extent of overlap is

$$\frac{ab}{a+b}.$$

Proof: Without loss of generality, assume that $b \geq a$. Place the origin at the beginning of the a range. Then the right limit of the b range can lie between 0 and $a+b$ with uniform probability. Thus, the expected overlap is simply given by

$$\frac{\int_0^a x dx + (b-a)a + \int_b^{a+b}(a+b-x)dx}{\int_0^{a+b} dx}$$

which evaluates to

$$\frac{ab}{a+b}$$

as was originally claimed.∎

Lemma 2. Given $k > 1$ constraints each of extent α, the expected range of overlap is

$$\frac{\alpha}{k}.$$

Proof: The proof is by induction. Clearly the $k = 1$ is trivial. The $k = 2$ case follows straightforwardly from Lemma 1. Now assume the induction hypothesis is true for $k-1$ and establish it for k. By Lemma 1, adding the k^{th} constraint yields an expected range of

$$\frac{\frac{\alpha}{k-1}\alpha}{\frac{\alpha}{k-1}+\alpha}$$

and this reduces to

$$\frac{\alpha}{k}$$

thereby establishing the induction hypothesis.∎

This second lemma implies that, in the three degree of freedom case being discussed here,

$$\left|\bigcap_{m=1}^{j-1} \mathbf{V}^{*}_{\epsilon,\ell,\gamma}\right| \leq \frac{\left|\mathbf{V}^{*}_{\epsilon,\ell,\gamma}\right|}{(j-1)^3}.$$

Thus, using equation (2), the expected number of interpretations is bounded by

$$\sum_{i=1}^{n} \prod_{j=2}^{k+1} \frac{n}{(j-1)^3} \frac{|\mathbf{V}^*_{i,\ell,\gamma}|}{v_T}.$$

If we let v^* denote this normalized volume

$$v^* = \frac{|\mathbf{V}^*_{i,\ell,\gamma}|}{v_T}$$

then the expected number of interpretations is bounded by

$$n\frac{[nv^*]^k}{(k!)^3}.$$

Since it is well known that

$$k! > \left(\frac{k}{e}\right)^k$$

this reduces to an expected upper bound of

$$n\left\lceil \frac{ne^3 v^*_{i,\ell,\gamma}}{k^3} \right\rceil^k. \tag{4}$$

This is clearly a much tighter bound than that of equation (3). Indeed, for fixed n, this behaves like

$$n\left\lceil ck^{-3} \right\rceil^k$$

and the k^{-3k} term is extremely important, since for fixed n this function is unimodal, rising to a peak for some value of k dependent on the values of the constants, and then dramatically decreasing as k is further increased. A sample plot is shown in Figure 3. In Figure 4, the log of the number of expected interpretations is plotted as a function of k, with each plot corresponding to a doubling of the constant of the previous plot.

It is worth noting that exactly this type of behavior was observed empirically by [Grimson and Lozano-Pérez 84]. In Figure 5, we plot the number of consistent interpretations as a function of the number of data points for an actual example. Note that the form of the graph is very similar to that of Figure 3. The fact that the number of interpretations does to asymptotically approach 1 is due to symmetries in the object which violate the assumptions of uniform distribution of faces in RC-space. Even so, the performance on real data closely matches that predicted by theoretical estimates.

Finally, it is clear that none of the above discussion was critically dependent on the fact that we were dealing with the two dimensional case, although it was easy to visualize in this case. Similar constraints in the three dimensional case will also result in a restriction of the swept volume of RC-space, appropriately defined. Here, the spatial components of the RC-space will be similarly bound. In particular,

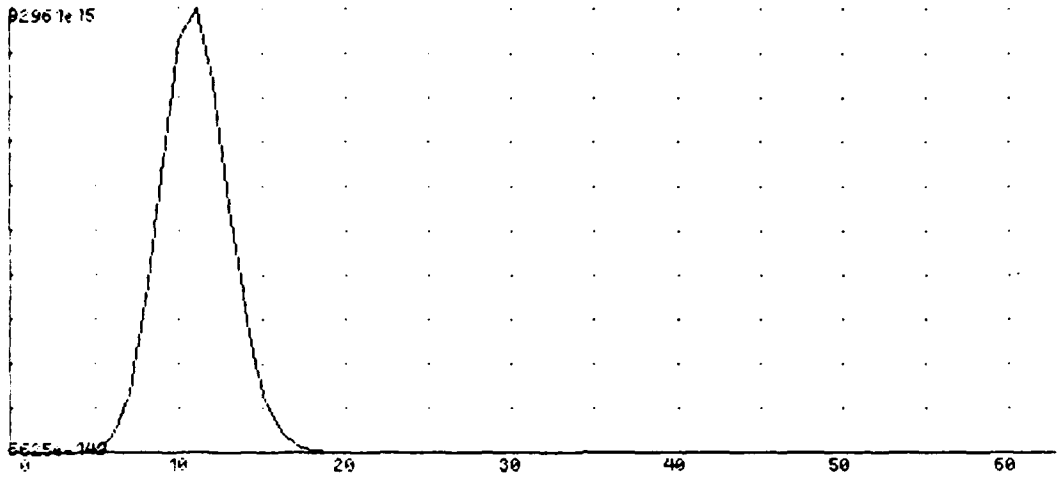


Figure 3. A sample plot of $n[ck^{3}]^{k}$ for fixed n and c and increasing k. The function is unimodal, peaking for some small value of k.

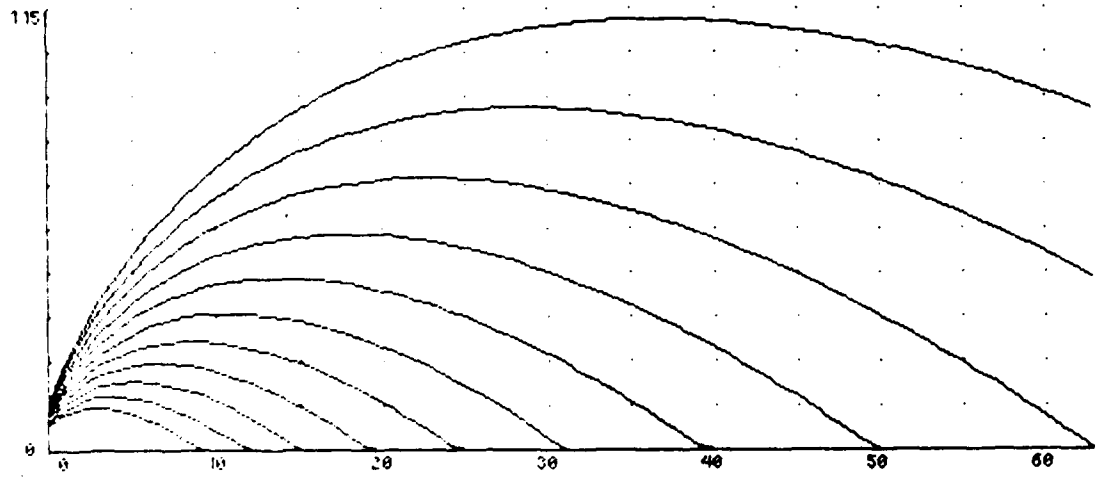


Figure 4. The logarithm of the expected number of interpretations is plotted as a function of k. Each graph corresponds to a doubling of the constant used in the previous plot.

if j is the number of independent constraints, defining the swept volume $|\mathbf{V}^{*}_{\epsilon,\ell,\tau}|$ then the expected bound on the number of interpretations for $k+1$ sensory points is

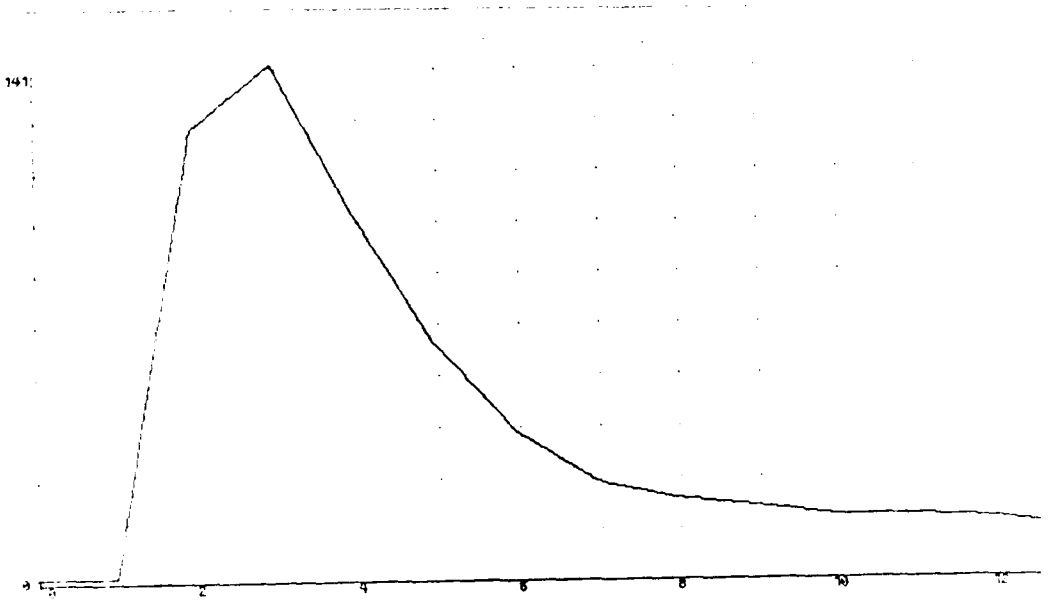


Figure 5: The number of consistent interpretations for a real example is plotted as a function of the number of sensory data points. The form of the graph should be compared to figure 3.

$$n^{\left\lfloor \frac{ne^j v^*_{\epsilon,\delta}}{k^j} \right\rfloor^k}. \tag{5}$$

For the two dimensional case, we have seen that $j = 3$ and for the three dimensional case, $j = 5$.

Note that in using Lemma 2 to establish these bounds, we implicitly assumed that the volumes of RC-space obtained by intersecting the swept volumes defined by the sensory constraints must be non-empty. This must hold true, of course, in the case of the correct interpretation, but for other cases, the intersected volume need not be empty. This implies that tighter bounds could be obtained by taking this into account, since in such cases, the expected overlap would be smaller than the expression derived in Lemma 2. We also note that while the expected volume defined by equation (5) tends to 0 with increasing k, this does not imply that the number of feasible interpretations also tends to 0. This is since the faces of the object are represented by δ functions in this space, and thus a continuous limit process such as equation (5) may still contain a δ function within its scope, even as the volume described by it shrinks to an infinitesimal. So long as the δ functions representing different faces are not arbitrarily close to one another, for example, under the assumption of uniform distribution, the asymptotic reduction of equation (5) with increasing k does imply a corresponding reduction in the number of interpretations, and equation (5) could be amended by adding a ceiling function *to the entire expression*.

Finally, we stress that to this point, we have not actually used the specific forms of the pairwise constraints, but only the fact that such constraints restrict the

ranges of parameters that describe relationships between faces of an object. Thus, provided we can bound the volume of RC-space spanned by such a constraint, the results derived above give us a means of measuring the efficiency of such constraints in restricting the set of feasible interpretations of the data, and the behavior of the bounds with increasing numbers of data points allow us to measure the efficiency of the constraints in restricting the search space.

4.2. Predictions of the Bounds

Several interesting factors can be derived from these bounds. We already have noted that this bound is a unimodal function in k. We first derive an expression for the value k_0 for which this maximum is achieved. Suppose we let

$$E(k, j, n, v^*_{I,S}) = n \left[\frac{n e^j v^*_{I,S}}{k^j} \right]^k$$

denote the number of expected interpretations for $k+1$ points of sensory information, where j is the number of dimensions of the RC-space that are explicitly constrained by the local constraints. By setting the derivative of the logarithm of E to zero, we find that the function E reaches a maximum for

$$k_0 = \{nv^*\}^{\frac{1}{j}}.$$

At this point, the bound on the expected number of feasible interpretations is simply

$$n \exp j \{nv^*\}^{\frac{1}{j}}$$

so that this places a bound on the maximum number of interpretations that must be considered at any point in the generation of the tree of interpretations.

Secondly, we can invert this bound to predict the number of data points needed, in general, to guarantee a unique determination of the object's orientation. In particular, given a value of E (e.g. $E \leq 1$) we can use Newton's method to solve for the number of data points needed to reduce the bound in equation (5) below this level. While numeric solution of this implicit equation for k will yield the best results for specific values of n, j and v, we can also derive an upper bound on the number of data points required to produce a unique interpretation. This is given by a solution for k of the following equation:

$$(k+1)\log n + jk + k \log v^* - jk \log k \leq 0$$

or

$$k \left[j \left(\frac{\log k - 1}{\log n} \right) - a - 1 \right] \geq 1$$

where we have replaced the variable v^* by the variable a through the relationship

$$v^* = n^a.$$

One method of ensuring that this expression is true is to require that both terms in the product on the left hand side are greater than 1. In particular, the second term then reduces to the inequality

$$\log \kappa \geq 1 + \binom{\alpha + 2}{j} \log n$$

and this yields, upon substitution, the bound

$$\kappa \geq e\, n^2 v^{*\,\frac{1}{j}}.$$

This is, of course, a weak bound, but it does provide us with an upper estimate for the number of sensory points needed, in general, to guarantee a unique interpretation of the data.

1.3. Degradation with noise

We can use the bounds derived above to measure the degradation in our recognition techniques with increasing sensor error. In particular, we can examine the number of points for which a maximum number of feasible interpretations occurs,

$$k_0 = n v^{*\,-\frac{1}{j}},$$

the maximum number of feasible interpretations at that point,

$$m = n \exp\left\{j\, n v^{*\,\frac{1}{j}}\right\},$$

and the weak upper bound on the number of points needed to guarantee, in general, a unique interpretation, up to partial symmetries of the object, of course,

$$\rho = e\, n^2 v^{*\,-\frac{1}{j}}.$$

The question of concern is how these parameters vary with increased noise in the sensory data. Clearly only v^*, the relative volume of R^n-space, will increase as the sensor error is increased. In particular, we examine the relative change in our parameters k_0, m, and ρ if we let ϵ denote the bound on sensor error (be it measuring position error or surface orientation error), it is straightforward to show, using the results of the previous section, that

$$\frac{\partial k_0}{\partial \epsilon} = -\frac{1}{j} \frac{\partial \log v^*}{\partial \epsilon}$$

$$\frac{\partial m}{\partial \epsilon} = k_0 \frac{\partial \log v^*}{\partial \epsilon}$$

$$\frac{\partial \rho}{\partial \epsilon} = -\frac{1}{j} \frac{\partial \log v^*}{\partial \epsilon}.$$

Thus, we see that the term

$$\frac{\partial \log v^*}{\partial \epsilon}$$

is the critical factor in determining the graceful degradation of the constraints. Provided this expression is small, the relative changes in these parameters are also small, and thus the combinatorics of the constraints degrades gracefully. Note that this graceful degradation improves as the number of independent constraints j increases.

5. Specific Bounds

We now turn to a careful consideration of the specific constraints that apply to our particular case, and derive explicit expected bounds on the number of interpretations consistent with $k+1$ sensory points, by estimating the volume parameter

$$V^*_{\ell,\epsilon,\gamma}.$$

5.1. The Two Dimensional Case

We will first consider the two-dimensional case of objects with three degrees of freedom relative to their model coordinates, and will then extend this to the general three dimensional case of objects having six degrees of positional freedom.

First, we will make the following assumptions about the object being sensed

1. *All the edges of the object have the same length ℓ.*
2. The diameter of the object is denoted by the constant D.

We make the following assumptions about the sensory information

1. The positions of contact with the object are known to within a circle of radius ϵ.
2. The normals to the sensed edges are known to within an angular error of γ.

Now suppose that we arbitrarily choose some face on which to place the first contact point (obviously we have no constraints on this anyway). Given a second sense point, the first question to consider is how many faces are consistent with the constraints between the two sensory points. To answer this, we first review what constraints are available from two sensory points.

1. We can measure (to within some error) the angle θ_0 between the normal of the first face and the normal of the second face. This is given by the combination of knowing the dot product between the two normals and the cross product of the two normals (which is the simplified version of the triple product constraint in the case of two dimensions). In fact, this measurement is accurate to within a range of $\pm\gamma$.
2. We can measure the length d of the contact vector $\mathbf{v}$ between the two sensed points (with a possible error of $\pm\epsilon$).

3. Finally, we can determine the angle between the contact vector v and the first face normal. To determine the range of possible error, we note that each endpoint of v lies within a circle of radius ϵ. It is straightforward to show, either algebraically or geometrically, that the maximum angular deviation of the true contact vector from its measured position is given by

$$\tan^{-1}\left(\frac{2\epsilon}{d}\right).$$

Our general technique in estimating bounds on the combinatorial efficiency of these constraints will be the following:

1. Construct a relative configuration space (RC-space) centered at the midpoint of the first face.
2. Bound the volume of this space in which the second face must lie.
3. Estimate the number of faces lying within this volume. To do this, we will assume that the faces are uniformly distributed over a volume of RC space within D of the midpoint of the first face. While this will generally be true over a large collection of objects, clearly for any single object this will not be completely valid. We will argue that the bounds derived under this assumption are *almost everywhere* bounds. That is, while there may be certain degenerate cases which violate the bounds, in general, over the class of possible objects, the bounds will almost always be true.
4. Extend the analysis to consider the combinatorial effect of k points on the number of faces assignable to the $k+1^{\text{st}}$ point.

We arbitrarily assign a face to the first sense point, and construct a relative configuration space about this face. The origin of the spatial dimensions of the RC-space lies at the midpoint of the face. The normal of this face defines the origin of the angular dimension of the RC-space. Thus, a face is represented in RC-space by the position of its midpoint relative to the midpoint of the base face, and by the angle of its normal relative to the normal of the base face.

Suppose that the base of the contact vector v lies at one of the endpoints of the first face. What positions could the endpoint of v take in RC-space? First, we know that the angle of this vector relative to the sensed normal to within a range $\tan^{-1}(2\epsilon/d)$ and we know that the variation in the sensed normal relative to the face normal is given by γ. Thus, the angle made by the contact vector relative to the face normal is defined by the range

$$\tan^{-1}\left(\frac{2\epsilon}{d}\right)+\gamma.$$

Thus the area of RC-space swept out by the endpoint of v is given by

$$\int_{\rho=d-\epsilon}^{d+\epsilon}\int_{\theta=\theta_0-\gamma-\tan^{-1}\frac{2\epsilon}{d}}^{\theta_0+\gamma+\tan^{-1}\frac{2\epsilon}{d}}\rho\,d\rho\,d\theta$$

which evaluates to

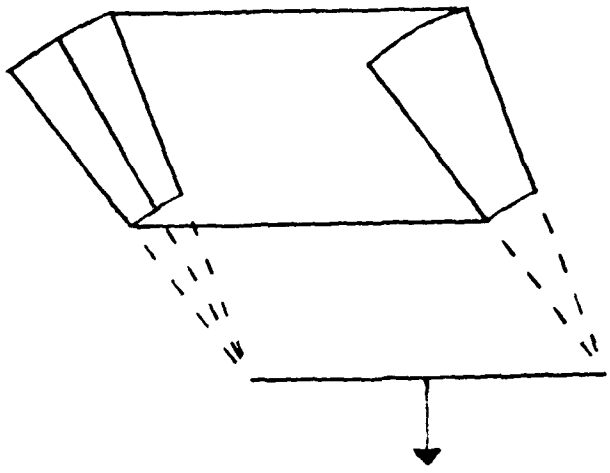


Figure 6 The swept volume of possible positions for a sensory point, *given* that the first sensory point lies on the base face.

$$4\left(\gamma + \tan^{-1}\frac{2\epsilon}{d}\right)\epsilon d.$$

The total area of RC-space swept out as the base point of v is moved across the base face is diagrammed in Figure 6.

In order to ease the what follows, we consider the problem of enscribing a rectangle, with sides aligned along the coordinate axes of the RC-space, about this volume. We let

$$\alpha = \gamma + \tan^{-1}\frac{2\epsilon}{d}.$$

We will assume that $\gamma \leq \frac{\pi}{4}$ and that we only consider contact points such that $d \geq 2\epsilon$. In this case, $0 \leq \alpha \leq \frac{\pi}{2}$. We may assume, without loss of generality, that θ_0 lies in the range $[0, \frac{\pi}{2}]$. In order to determine bounds on the dimensions of the enscribing rectangle, we consider three case.

Consider first the case of $\theta_0 > \frac{\pi}{2} + \alpha$. In this case, one can show that

$$\begin{aligned} y_{\min} &= (d-\epsilon)\sin(\theta_0+\alpha) \\ y_{\max} &= (d+\epsilon)\sin(\theta_0-\alpha) \\ x_{\min} &= (d+\epsilon)\cos(\theta_0+\alpha) \\ x_{\max} &= (d-\epsilon)\cos(\theta_0-\alpha)+\ell. \end{aligned}$$

Similarly, if $\theta_0 < \frac{\pi}{2} - \alpha$ then

$$\begin{aligned} y_{\min} &= (d-\epsilon)\sin(\theta_0-\alpha) \\ y_{\max} &= (d+\epsilon)\sin(\theta_0+\alpha) \\ x_{\min} &= (d-\epsilon)\cos(\theta_0+\alpha) \\ x_{\max} &= (d+\epsilon)\cos(\theta_0-\alpha)+\ell, \end{aligned}$$

and if $\frac{\pi}{2} - \alpha < \theta_0 < \frac{\pi}{2} + \alpha$ then

$$\begin{aligned} y_{\min} &= (d - \epsilon)\min\{\sin(\theta_0 - \alpha), \sin(\theta_0 + \alpha)\} \\ y_{\max} &= (d + \epsilon) \\ x_{\min} &= (d - \epsilon)\cos(\theta_0 + \alpha) \\ x_{\max} &= (d + \epsilon)\cos(\theta_0 - \alpha) + \ell. \end{aligned}$$

Thus, for example, the range in y is given in these three cases by

$$\begin{aligned} \Delta y &= -2d\cos\theta_0\sin\alpha + 2\epsilon\sin\theta_0\cos\alpha && \text{Case 1} \\ &= 2d\cos\theta_0\sin\alpha + 2\epsilon\sin\theta_0\cos\alpha && \text{Case 2} \\ &= (d + \epsilon) - (d - \epsilon)\min\{\sin(\theta_0 - \alpha), \sin(\theta_0 + \alpha)\}. && \text{Case 3} \end{aligned}$$

We want to consider the maximum extent of this range. Applying standard minimization techniques, we find that the maximum value for both Case 1 and Case 2 are given by

$$\Delta y = 2\sqrt{d^2\sin^2\alpha + \epsilon^2\cos^2\alpha}$$

and for Case 3, the maximum occurs at

$$2\left(d\sin^2\alpha + \epsilon\cos^2\alpha\right).$$

Moreover, the extremum observed in Case 1 and Case 2 is always greater than that of Case 3, so we can bound the y component of the enscribing rectangle by

$$\Delta y = 2\sqrt{d^2\sin^2\alpha + \epsilon^2\cos^2\alpha}.$$

In a similar manner, we find that we can bound the x component of the enscribing rectangle by

$$\Delta x = \ell + 2\sqrt{d^2\sin^2\alpha + \epsilon^2\cos^2\alpha}.$$

This gives us a rectangular bound on the area of RC-space in which the second sensory point can lie, given that the first sensory point lies somewhere on the base face. Since a face is described in RC-space by the position and orientation of its midpoint, relative to the base face, we need to expand this volume to cover the possible positions of the midpoints of the face. Clearly the most straightforward way to do this is to expand this rectangle by $\frac{\ell}{2}$ on all sides, since a sensed point must lie within that distance of the midpoint of the face. This yields an area in RC-space of dimensions

$$2\ell + 2\sqrt{d^2\sin^2\alpha + \epsilon^2\cos^2\alpha}$$

by

$$\ell + 2\sqrt{d^2\sin^2\alpha + \epsilon^2\cos^2\alpha}$$

in which the midpoint of a face must lie in order to be consistent with the sensory information.

Finally, we must account for the orientation of the face. Clearly, the angular component of RC-space has extent 2π and the range of possible values for the orientation of the second face relative to the first is given by 2γ.

This defines a volume in RC-space in which a face must lie in order to be consistent with the sensory information. In order to make useful estimates concerning this volume, we need to normalize this volume. To do so, we will make the assumption that the faces have equal probability of lying anywhere within this RC-space (which spans an area πD^2 in the spatial dimensions and 2π in the angular dimension). Clearly this assumption is not necessarily valid for any particular object, although when averaged over all possible orientations of the object, it becomes more accurate. In this case, the normalized volume of RC-space in which a face must lie is bounded above by

$$V_2 = \frac{2}{\pi} s_1 \times s_2$$

where

$$s_1 = \left[\frac{2\ell}{\sqrt{\pi D}} + 2\sqrt{\frac{\sin^2\alpha}{\pi} + \left(\frac{\epsilon}{\sqrt{\pi D}}\right)^2 \cos^2\alpha} \right]$$

$$s_2 = \left[\frac{\ell}{\sqrt{\pi D}} + 2\sqrt{\frac{\sin^2\alpha}{\pi} + \left(\frac{\epsilon}{\sqrt{\pi D}}\right)^2 \cos^2\alpha} \right].$$

The number of possible faces lying within this volume is given by

$$nV_2.$$

Note that, as expected, V_2 is a dimensionless quantity, depending only of ratios of parameters of the polygonal object. Also, we can restrict our computation to cases where $d \geq 2\epsilon$ so that α can be bounded above by $\gamma + \frac{\pi}{4}$.

This normalized volume leads to a straightforward bound on the number of interpretations consistent with $k+1$ sense points, namely

$$n[nV_2]^k.$$

Of course, this ignores much of the combinatorial constraint available to the technique, since it comes from simply using the constraints between the first sense point and all additional sense points but ignores the fact that the k^{th} face is constrained by all $k-1$ previous sense points, and not just the first. Nonetheless, under the assumptions made previously, this bound represents a worst case bound, in which the inclusion of additional sense points does not reduce the normalized volume of RC-space specified by the constraints between the first and the current sense points.

We note that for fixed n, the expression for the constant nV_2 actually defines a relationship between the two error ranges as a function of the object parameters.

Thus, it describes the tradeoff between the error in position measurements and normal measurements needed to guarantee convergence of the combinatorics with increasing sensed point. Note that product relationship between these errors makes sense, since an increase in errors in position should require a decrease in errors in normals in order to maintain the same number of interpretations, and vice versa.

This bound is actually quite weak, however, since it loses a great deal of the combinatorial power of combining different constraints. Consider the $k+1^{st}$ point. Not only should the pairwise constraints between the first sensed point and the $k+1^{st}$ point restrict the volume of RC-space in which the $k+1^{st}$ face can lie, but so should the pairwise constraints between the second point and the $k+1^{st}$ point and so on. While all of these volumes could be identical on average we expect the volumes to only partially overlap. Thus, the volume for the $k+1^{st}$ face should be the intersection of all of these volumes. A bound on this intersected volume is given by Lemmas 1 and 2. Thus, for example, the angular component for the $k+1^{st}$ is reduced by a factor of k.

We would expect a similar effect for the position components of RC-space, but the argument here is slightly more complicated. If it were the case that the error ellipse rectangles derived above were always aligned with the coordinate axes, then an identical argument would hold. Since the faces will be a different orientations with respect to the coordinate frame of the RC-space, however, when finding the area of intersection of two such constraints, one cannot straightforwardly decompose the two components of the rectangle. To get around this problem, we can consider enscribing the enscribed rectangle within a larger square, such that the initial enscribed rectangle will fall within the square, independent of its orientation. Clearly, the side of the enscribed square is given by the diagonal of the initial enscribed rectangle, and this is given by

$$\beta_2 = \sqrt{s_1^2 + s_2^2}.$$

Using this bound on the ranges within each of the spatial dimensions of the normalized RC-space, we can use the previous lemmas to obtain bounds on the expected volume of RC-space for the $k+1^{st}$ face. In particular, the expected volume is bounded by

$$\frac{\gamma}{\pi k} \frac{\beta_2^2}{k^2}.$$

Piecing this together over all the points, we find that the expected number of interpretations consistent with $k+1$ sensory points is bounded above by

$$n^{\left\lceil \frac{e\gamma\beta_2^2 e^3}{\pi k^3} \right\rceil^k}.$$

5.2. The Three Dimensional Case

We can expand the derivation of the previous section to deal with the three dimensional case. Here, the faces are squares of side ℓ, rather than one dimensional

edges. It is straightforward to show that the positional aspects of the derivation can be decomposed into x and y components, both yielding normalized ranges of size s_1, while the z component has a normalized range given by s_2. Only one of the two rotational parameters can be constrained by the measurements between the two unit surface normals, so that the normalized volume in this case is given by

$$V_3 = \frac{\gamma}{\pi} s_1^2 \times s_2.$$

An expected bound can be derived in a manner similar to the two dimensional case, where now the diagonal of the three dimensional cube is given by

$$\beta_3 = \sqrt{2s_1^2 + s_2^2}.$$

Here the expected volume is

$$\frac{\gamma}{k\pi} \frac{\beta^3}{k^3}.$$

We have yet to apply the triple product constraint [Grimson and Lozano-Pérez 1984], however. While this constraint has no application to pairs of points, for triples of points it does apply. In particular the following case holds. Consider three independent unit normals $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$, and assume we know the dot products between each of them, denoted by ϵ_{ij}, as well as the sign of the triple product $[\mathbf{n}_1\mathbf{n}_2\mathbf{n}_3]$. Then

$$\mathbf{n}_2 = \epsilon_{12}\mathbf{n}_1 + \sqrt{1-\epsilon_{12}^2}\mathbf{u}$$

where $\mathbf{u}$ is a unit vector such that $\mathbf{u} \cdot \mathbf{n}_1 = 0$. Similarly

$$\mathbf{n}_3 = \epsilon_{13}\mathbf{n}_1 + \sqrt{1-\epsilon_{13}^2}\mathbf{w}$$

where $\mathbf{w}$ is a unit vector such that $\mathbf{w} \cdot \mathbf{n}_1 = 0$. Now the dot product between these vectors is known by measurement,

$$\begin{aligned}\mathbf{n}_2 \cdot \mathbf{n}_3 &= \epsilon_{12}\epsilon_{13} + \sqrt{1-\epsilon_{12}^2}\sqrt{1-\epsilon_{13}^2}\,\mathbf{u} \cdot \mathbf{w} \\ &= \epsilon_{23}.\end{aligned}$$

Thus, the cosine of the angle between $\mathbf{u}$ and $\mathbf{w}$ is determined, say $\mathbf{w} \cdot \mathbf{u} = \cos\theta$. Evaluating the triple product gives

$$[\mathbf{n}_1\mathbf{n}_2\mathbf{n}_3] = \sqrt{1-\epsilon_{12}^2}\sqrt{1-\epsilon_{13}^2}\sin\theta$$

and thus the angle θ is determined. This implies that while there is a free degree of freedom between $\mathbf{n}_1$ and $\mathbf{n}_2$, corresponding to a rotation about $\mathbf{n}_1$, the attitude of $\mathbf{n}_3$ is completely determined from the measured dot products and triple product.

This implies that while the expected volume for the second point is given by

$$\frac{\gamma}{k\pi} \frac{\beta^3}{k^3}$$

for additional points, the second rotational degree of freedom is also restricted and the expected volume is given by

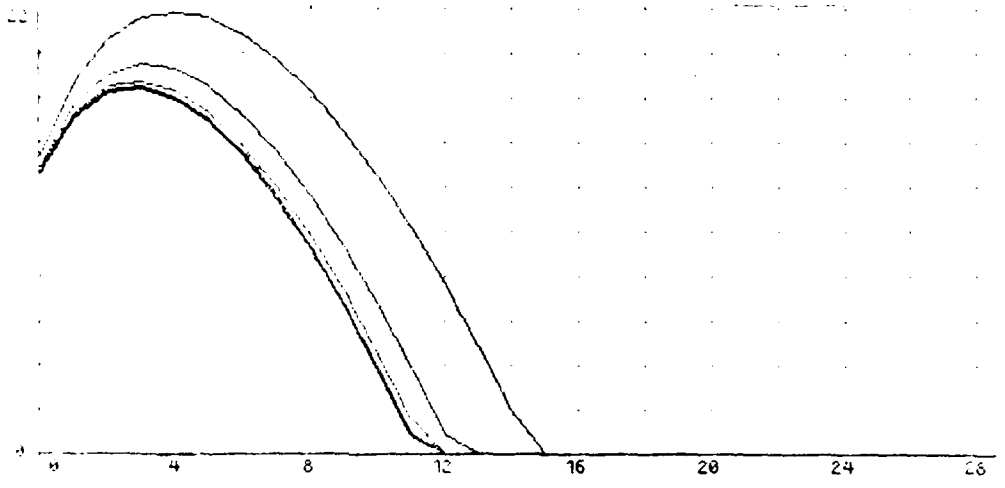


Figure 7. The logarithm of the expected number of interpretations as a function of the number of sensory points. Each graph illustrates the effects of doubling the amount of error in the surface positions, with all other parameters fixed.

$$\left(\frac{2}{k\pi}\right)^2 \frac{\beta^3}{k^3}.$$

Thus, the expected number of interpretations for $k+1$ sensory points is bounded above by

$$n^{\pi\left[\frac{n\gamma_1\beta_3^2 e^5}{\pi k^5}\right]^k}.$$

5.3. Degradation with Noise

Having derived specific bounds on the effective pruning of the local constraints, it is easy to see how the presence of sensor error affects the expected number of consistent hypotheses. To demonstrate the graceful degradation of the constraints, we have plotted the logarithm of the number of expected feasible interpretations as a function of the number of sensory points, for several different error conditions.

In Figure 7, we plot the number of interpretations while varying the amount of error in measuring surface positions. Each plot corresponds to a doubling of this positional error from the previous one.

In Figure 8, we plot the number of interpretations while varying the amount of error in measuring surface orientation. Each plot corresponds to a doubling of this angular error from the previous one. In both cases, we see that the combinatorics generally decrease in a graceful manner.

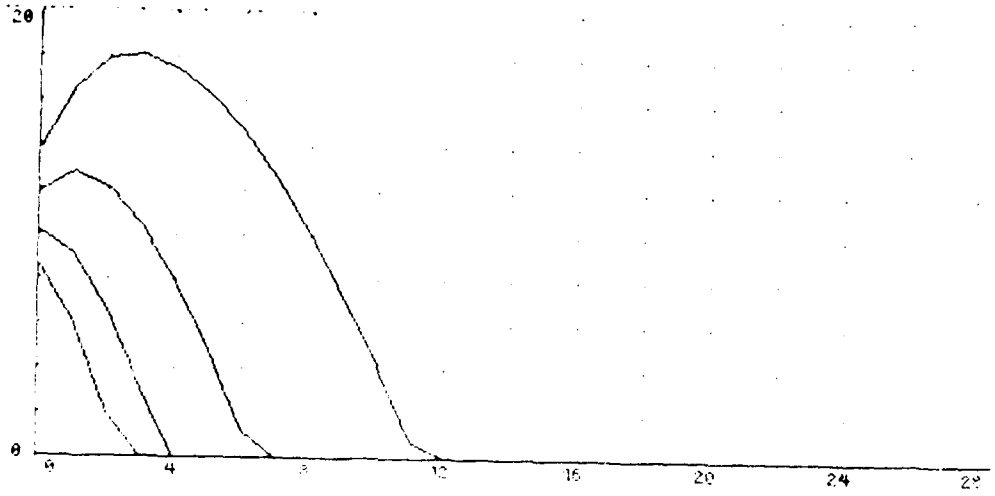


Figure 8. The logarithm of the expected number of interpretations as a function of the number of sensory points. Each graph illustrates the effects of doubling the amount of error in measuring surface orientation, with all other parameters fixed.

6. Application of the Theory

We began our investigation of the problem of model-based recognition and localization by establishing a set of criteria that should be satisfied by constraints between sensory data and model elements. In particular, we argued that the constraints should be coordinate frame independent, simple, sensor independent, combinatorially powerful and degrade gracefully with error. We have spent the bulk of this paper deriving such a set of constraints and establishing on theoretical grounds that these criteria are satisfied. This theoretical basis was an outgrowth of earlier work [Grimson and Lozano-Pérez 1984] in which an isomorphic set of constraints was proposed and tested.

As further support for the theoretical results reported here, we briefly review some of the results of testing that algorithm. In particular, a large set of simulations were run on a series of test objects, for varying types of error conditions. The following table summarizes some of these simulations, described in more detail in [Grimson and Lozano-Pérez 84], for the objects illustrated in Figure 9.

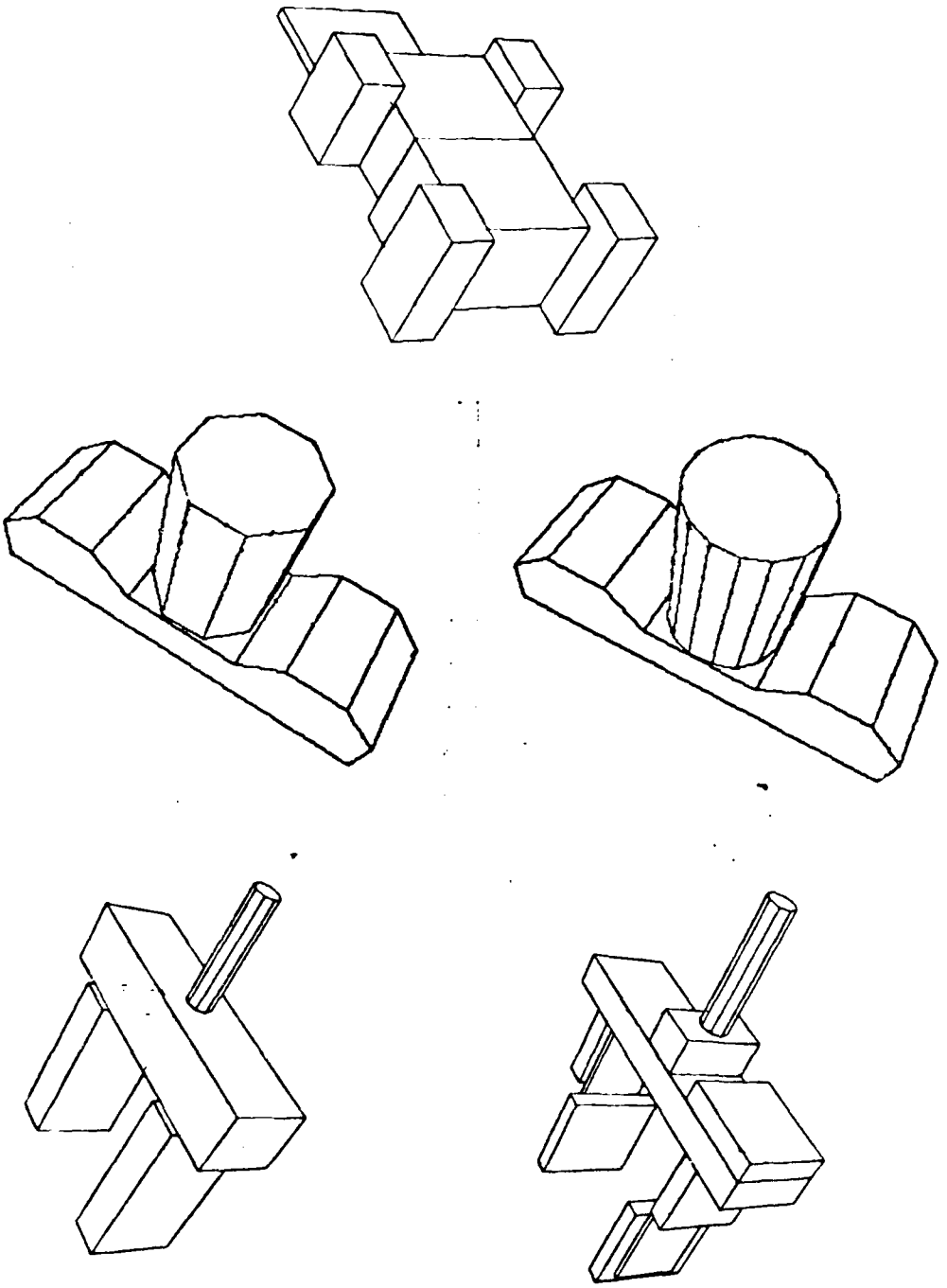


Figure 9. 3D Test Models

Table 1 No. of Interpretations After Local Pruning						
Object	Normal	Dist	Min	50th	95th	Faces
Housing	$\pi/15$	.01	1	1	2	40
		.05	1	1	4	40
		.10	1	4	17	40
	$\pi/10$	.01	1	1	3	40
		.05	1	2	9	40
		.10	1	8	43	40
	$\pi/8$	.01	1	1	4	40
		.05	1	3	16	40
		.10	1	12	96	40
Simple Hand	$\pi/12$	.01	2	2	8	28
		.05	2	4	12	28
	$\pi/10$	.01	2	2	16	28
		.05	2	4	16	28
	$\pi/8$	.01	2	4	32	28
		.05	2	4	32	28
Complex Hand	$\pi/12$	.01	1	4	24	64
		.05	1	10	64	64
	$\pi/10$	.01	1	4	48	64
		.05	1	15	94	64
	$\pi/8$	.01	2	12	58	64
		.05	2	18	136	64
8 Cylinder	$\pi/10$	.08	2	4	12	21
		.16	2	4	29	21
	$\pi/8$	.08	2	8	33	21
		.16	2	16	106	21
	$\pi/7$	.08	2	12	68	21
		.16	4	40	336	21
16 Cylinder	$\pi/10$	.08	4	36	216	29
		.16	4	60	504	29
	$\pi/8$	.08	4	84	642	29

The table above illustrates statistics in the performance of the localization process. Each row lists parameters of a histogram of the number of interpretations consistent with the local constraints, based on 100 trials with the object randomly oriented with 6 degrees of freedom. In each case, 12 sensory data points were used. In the table, the *normal* column lists the radius of the error cone about the measured surface normal; the *dist* column lists the error range of the distance sensing; the *min* column list the minimum number of interpretations observed; the *50th* column lists the median point of the set of simulations; the *95th* column lists the 95^{th} percentile of the set of simulations; and the *faces* column lists the number of faces in the model.

These simulations are very encouraging, and support the theoretical results made here. In particular, they demonstrate the dramatic reduction in the number of feasible interpretations of very few, sparse, noisy data points. The gradual degradation with the addition of noise can be seen by comparing appropriate rows in the table. In addition to the simulations summarized above, the technique has also been applied to several different types of real data, including sonar range data, binary images and edges detected from grey-level images.

7. Discussion

We have shown that local coordinate frame constraints between sparse, noisy sensory data are very effective in reducing the recognition and localization problem. Several points of interest arise from the analysis.

First, in order to establish general bounds on the expected number of interpretations, we made the assumption that, by averaging over a number of objects, we could treat the distribution of faces in relative configuration space as uniform. Of course, many objects will violate this assumption, and we must ask how valid is the assumption and the effect of relaxing the assumption on the derived bounds.

A simple means of examining the first question is to construct a specific example. Consider the simple, two-dimensional object illustrated in Figure 10(a). The spatial distribution of midpoints of faces of this object in relative configuration space is illustrated in Figure 10(b). Several points are noticeable in this diagram. The first is that the relative midpoints of the faces do tend to uniformly fill the disk of possible positions. There are some clusters, however, and these generally correspond to partial symmetries in the object. This suggests that one set of objects that will badly violate our assumption of uniform distribution is regular polyhedra. For example, a cube, because of its 4-fold symmetry, will have a face configuration consisting of 4 clusters of 4 indistinguishable points, and thus the number of consistent hypotheses *should be higher* than predicted by the combinatorial bounds. This is in fact observed in simulations reported in Grimson and Lozano-Pérez [8], and can be seen in the rows of Table I corresponding to the *3-cylinder* and the *10-cylinder*.

Note also that while the spatial components of the RC space are reasonably uniformly distributed for the object in Figure 10, since all the angles between faces are distinct, the rotational component of RC-space will only contain points corresponding to $\theta = 0$ (see Figure 11). That is, the spatial components of RC-space basically treat the configuration differently, while the rotational components do not. On the other hand, a regular polygon would have rotational RC-space points that are uniformly distributed, while the spatial components are strongly clustered. This suggest that in cases where only some (say p) of the components of RC space satisfy the criteria of uniform distribution, the expected number of consistent hypotheses is bounded by

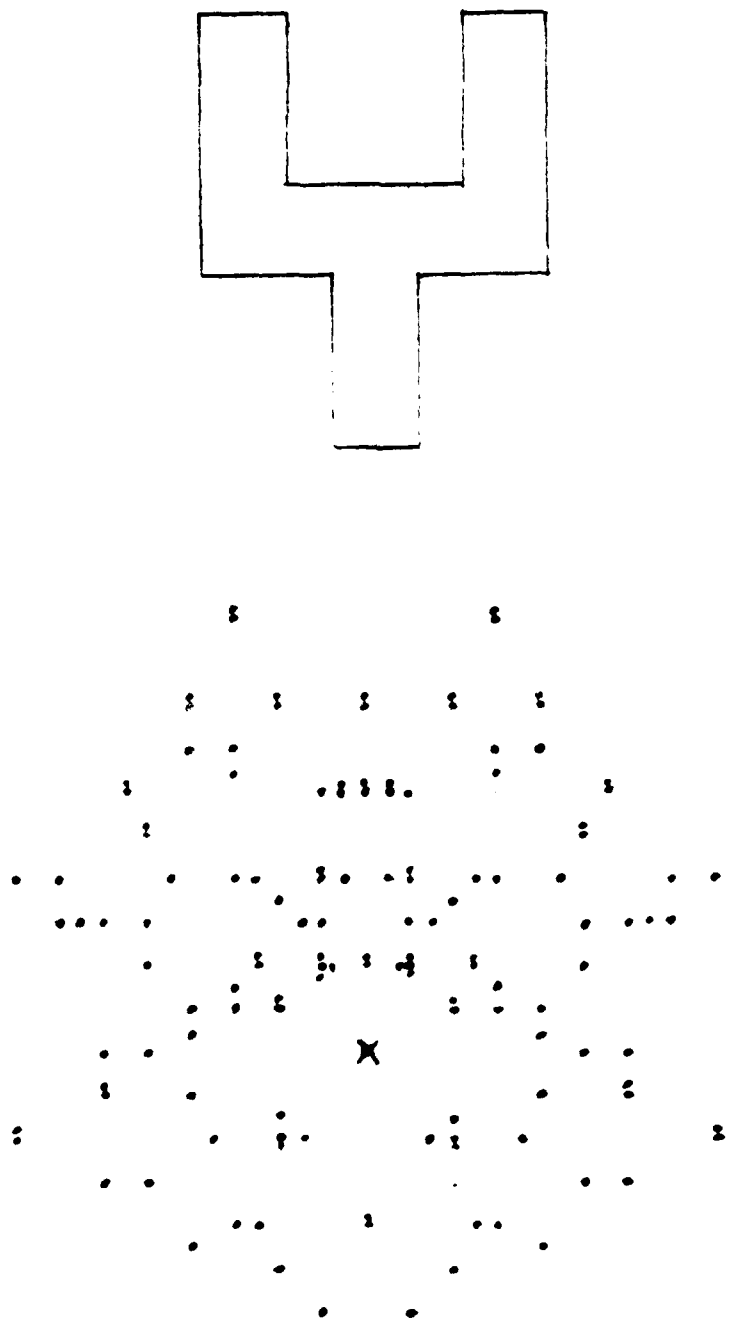


Figure 10. A sample object and its distribution of faces in relative configuration space.

$$n\left[\frac{ne^{j}\,V^{*}_{obj}}{k^{j}}\right]^{k}.$$

Thus, for example, in the case of the object illustrated in Figure 10, a more accurate bound would be given by using $j = 2$ rather than the more general $j = 3$.

Finally, the key results of the analysis presented here are in large part independent of the specific constraints used to restrict the search space. This suggests that a similar analysis using a swept volumes in a relative configuration space may be applicable to other cases of consistent interpretation of noisy, measured data. For example, extensions of the technique to curved surfaces and to other types of local sensory measurements remain interesting questions for future research.

Acknowledgments

Tomás Lozano-Pérez was critical to the development of this work, both in terms of the underlying recognition technique and in terms of valuable comments and criticisms of the material presented here. Matt Mason scrutinized versions of the manuscript with extreme care, and provided many valuable suggestions and criticisms. John Hollerbach and Berthold Horn provided valuable comments and criticisms at various stages.

References

Ballard, D. H. 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* 13(2):111–122.

Bausch and Lomb. 1978. Bausch and Lomb Omnicon Pattern Analysis System. Analytic Systems Division brochure. Rochester, New York: Bausch and Lomb.

Bolles, R. C., and Cain, R. A. 1982. Recognizing and locating partially visible objects: The Local-Feature-Focus method. *Int. J. Robotics Res.* 1(3):57–82.

Bolles, R. C., Horaud, P., and Hannah, M. J. 1983. 3DPO: A three-dimensional part orientation system. Paper delivered at First International Symposium of Robotics Research. Bretton Woods, N.H.

Brooks, R. 1981. Symbolic reasoning among 3-dimensional models and 2-dimensional images. *Artificial Intell.* 17:285–349.

Faugeras, O. D., and Hebert, M. 1983 (Aug. Karlsruhe, W. Germany). A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces. *Proc. Eighth Int. Joint Conf. Artificial Intell.* Los Altos: William Kaufmann, pp. 996–1002.

Gaston, P. C., and Lozano-Pérez, T. 1984. Tactile recognition and localization using object models: The case of polyhedra on a plane. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-6(3):257–265.

Gleason, G., and Agin, G. J. 1979 (Mar). A modular vision system for sensor-controlled manipulation and inspection. *Proc. Ninth Int. Symp. Industrial Robots*. Dearborn, Mich.: Society of Manufacturing Engineers, pp. 57–70.

Grimson, W. E. L., and Lozano-Pérez, T. 1984. Model-based recognition and localization from sparse range or tactile data. *Int. J. Robotics Res.* 3(3):3–35.

Holland, S. W. 1976 (Feb.) A programmable computer vision system based on spatial relationships. General Motors Publ. GMR-2078. Detroit: General Motors.

Horn, B. K. P., and Ikeuchi, K. 1983. Picking parts out of a bin. AIM-746. Cambridge, Mass.:Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Horn, B. K. P. 1983. Extended Gaussian images. AIM-740. Cambridge, Mass.:Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Ikeuchi, K. 1983. Determining attitude of object from needle map using extended gaussian image. AIM-714. Cambridge, Mass.:Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Marr, D. 1982. *Vision*. San Francisco:W. H. Freeman and Company.

Marr, D., and Nishihara, H. K. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond.* B200:269–294.

Machine Intelligence Corporation. 1980. Model VX-100 machine vision system. Product description. Mountain View, Calif.: Machine Intelligence Corporation.

Nevatia, R. 1974. Structured descriptors of complex curved objects for recognition and visual memory. Ph.D. thesis, Stanford University. AIM 250. Stanford, Calif.: Stanford University Artificial Intelligence Laboratory.

Nevatia, R., and Binford, T. O. 1977. Description and recognition of curved objects. *Artificial Intell.* 8:77–98.

Perkins, W. A. 1978. A model-based vision system for industrial parts. *IEEE Trans. Comput.* C-27:126–143.

Reinhold, A. G., and VanderBrug, G. J. 1980. Robot vision for industry: the autovision system. *Robotics Age*, Fall, pp. 22–28.

Stockman, G., and Esteva, J. C. 1984. Use of geometrical constraints and clustering to determine 3D object pose. TR84-002. East Lansing, Mich.:Michigan State University Department of Computer Science.

Sugihara, K. 1979. Range-data analysis guided by a junction dictionary. *Artificial Intell.* 12:41–69.

Tsuji, S., and Nakamura, A. 1975 (Aug., Cambridge, Mass.). Recognition of an object in a stack of industrial parts. *Proc. Fourth Int. Joint Conf. Artificial Intell.* Los Altos: William Kaufmann, pp. 811–818.